P

NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY

Faculty of Health, Natural Resources and Applied Sciences

School of Natural and Applied Sciences

Department of Mathematics, Statistics and Actuarial Science

13 Jackson Kaujeua Street
Private Bag 13388
Windhoek
NAMIBIA

T: +264 61 207 2913
E: msas@nust.na
W: www.nust.na

QUALIFICATION: Bachelor of Science Honours in Applied Statistics	
QUALIFICATION CODE: 08BSHS	**LEVEL:** 8
COURSE CODE: MVA802S	**COURSE NAME:** MULTIVARIATE ANALYSIS
SESSION: NOVEMBER 2024	**PAPER:** THEORY
DURATION: 3 HOURS	**MARKS:** 100

FIRST OPPORTUNITY EXAMINATION QUESTION PAPER	
EXAMINER	Dr D. B. GEMECHU
MODERATOR:	Prof L. PAZVAKAWAMBWA

INSTRUCTIONS

1. There are 6 questions, answer ALL the questions by showing all the necessary steps.
2. Write clearly and neatly.
3. Number the answers clearly.
4. Round your answers to at least four decimal places, if applicable.

PERMISSIBLE MATERIALS

1. Nonprogrammable scientific calculators with no cover.

THIS QUESTION PAPER CONSISTS OF 5 PAGES (Including this front page)

ATTACHMENTS

Two statistical distribution tables (z-and F-distribution tables)

Question 1 [12 Marks]

1.1. Define multivariate statistical analysis and state three objectives of multivariate analysis. **[4]**

1.2. Consider a random vector $\boldsymbol{y} = (Y_1, Y_2, Y_3)'$ with mean vector $\boldsymbol{\mu} = (3, 2, 1)'$ and covariance matrix $\boldsymbol{\Sigma}$. The eigenvalues of $\boldsymbol{\Sigma}$ are $\lambda_1 = 12, \lambda_2 = 6,$ and $\lambda_3 = 2$ and the corresponding eigenvectors are

$$\boldsymbol{e}_1 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, \boldsymbol{e}_2 = \begin{pmatrix} 2/\sqrt{6} \\ -1/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}, \boldsymbol{e}_3 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

Compute the following based on the information provided.

1.2.1. $|\boldsymbol{\Sigma}|$ **[1]**

1.2.2. trace (Σ) **[1]**

1.2.3. $Var(\boldsymbol{e}_1'\boldsymbol{y})$ **[2]**

1.2.4. $\boldsymbol{\Sigma}$ **[4]**

Question 2 [24 Marks]

2. Let $\boldsymbol{x} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & 0 \\ \rho^2 & 0 & 1 \end{pmatrix}$.

2.1. Drive the conditional distribution of (x_1, x_2) given x_3. **[8]**

2.2. If we define a new random variable $y = x_1 + \frac{1}{3}x_2 - \frac{1}{2}x_3$ and the value of $\rho = 0.5$, then:

2.2.1. derive the distribution of y and compute $P(y < -2)$ **[8]**

2.2.2. derive the joint distribution of x_3 and y. Are they independently distributed? Provide explanation for your answer. **[8]**

Question 3 [10 Marks]

3. Let the observation vector be partitioned into two subsectors denoted by $\boldsymbol{y}$ and $\boldsymbol{x}$, where $\boldsymbol{y}$ is $p \times 1$ and **x** is $q \times 1$. Assume that $\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix} \sim N_{q+p}\left[\begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}\right]$. Use this information to state and prove the independency properties of multivariate normal distribution. **[10]**

Question 4 [13 Marks]

4. Morphology is the branch of biology that deals with the form (structure) of living organisms. An expert measures the length (in cm) and weight (in hundreds of grams) of 20 adult birds from the same species, but from two different sub-species (10 birds in each sub-species). The data can be seen in the following figure, where the points are marked differently to distinguish observations from sub-species 1 and sub-species 2. From the multivariate data, we have the sample mean for Variety 1 as $\bar{\boldsymbol{y}}_1 = (45.4,\ 8.01)'$ and for Variety 2 $\bar{\boldsymbol{y}}_2 = (43.0, 10.06)'$ and pooled sample covariance matrix, $\boldsymbol{S}_{\mathrm{p}} = \begin{pmatrix} 3.578 & 2.053 \\ 2.053 & 2.002 \end{pmatrix}$

Assuming that the observations are bivariate and follow multivariate normal distributions $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, for $i = 1\ and\ 2$, Conduct a test if there is any significant difference between the vector of expected mean measurements of the two species at 5% level of significance. Your answer should include the following:

4.1. State the null and alternative hypothesis to be tested **[1]**

4.2. State the test statistics to be used and its corresponding distribution **[2]**

4.3. State the decision (rejection) rule and compute the tabulated value using an appropriate statistical table **[3]**

4.4. Compute the test statistics and write up your decision and conclusion **[7]**

Question 5 [25 Marks]

5. A veterinary scientist measured y_1 = Wing length (in cm) and y_2 = Back length (in cm) for a sample of n = 20 chickens. From this sample data we have sample mean vector $\bar{\boldsymbol{y}} = (4, 9)$, sample variance covariance matrix $\boldsymbol{S} = \begin{pmatrix} 9 & 16 \\ 16 & 64 \end{pmatrix}$, and the eigenvalue ($\lambda$) and eigenvector ($\boldsymbol{e}$) of $\boldsymbol{S}$ are

$$\lambda_1 = 68.315,\ \boldsymbol{e}_1 = \begin{pmatrix} 0.2604 \\ 0.9655 \end{pmatrix};\ \lambda_2 = 4.684,\ \boldsymbol{e}_2 = \begin{pmatrix} 0.9655 \\ -0.2604 \end{pmatrix}$$

Answer the following questions assuming that the sample were originated form a bivariate normal $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with unknown $\boldsymbol{\mu}$ and unknown $\boldsymbol{\Sigma}$.

5.1. Test the hypothesis $H_0\colon \boldsymbol{\mu} = (8,\ 12)'\ vs\ H_1\colon \boldsymbol{\mu} \neq (8,\ 12)'$ at 5% level of significance.

Your solution should include the following:

5.1.1. State the test statistics to be used and its corresponding distribution **[1]**

5.1.2. State the decision (rejection) rule and compute the tabulated value using an appropriate statistical table **[2]**

5.1.3. Compute the test statistics and write up your decision and conclusion **[7]**

5.2. Construct a 95% confidence ellipse for $\boldsymbol{\mu}$. **Hint:** Compute the major and minor axis of ellipse. **[8]**

5.3. Construct a 95%-T^2interval for a linear combination $\mu_1 - 0.5\mu_2$ **[7]**

Question 6 [16 Marks]

6. Heavy metals in river sediments are a significant concern due to their toxicological impacts on aquatic ecosystems and human health. Understanding the distribution and concentration levels of these metals across different rivers streams is essential for effective environmental management. Victor et al. (2024) measured heavy metals [Iron (FE), nickel (Ni), zinc (Zn), copper (Cu), manganese (Mn), and lead (Pd)] concentrations across three different river streams in Namibia- upper, middle and lower. A principal component analysis (PCA) was performed on a dataset the software output of the analysis of the data is given below. Answer the following questions based on these results

   6.1. Based on the correlation matrix provided, it the data suitable for PCA? **[2]**

   6.2. Discuss and interpret Bartlett's Test of Sphericity. **[3]**

   6.3. Test for sampling adequacy **[2]**

   6.4. Show that the value of $\lambda_6 = 0.131$ **[2]**

   6.5. What is the percentage of the total standardized variation attributed to the first principal component? **[2]**

   6.6. Using the results of the principal components analysis, draw a scree plot. How many principal components do you recommend to retain based on the scree plot? **[3]**

   6.7. Give a formula for computing the scores of the first principal component. Give a brief interpretation of the first principal component. **[2]**

######R-software output#######
##correlation matrix

	Fe	Mn	Ni	Pb	U	Zn
Fe	1.000	0.608	0.486	0.318	0.458	0.819
Mn	0.608	1.000	0.328	0.368	0.471	0.418
Ni	0.486	0.328	1.000	-0.006	0.147	0.313
Pb	0.318	0.368	-0.006	1.000	0.527	0.300
U	0.458	0.471	0.147	0.527	1.000	0.363
Zn	0.819	0.418	0.313	0.300	0.363	1.000

Bartlett's test for sphericity
$chisq
[1] 252.0737

$p.value
[1] 4.642919e-45

$df
[1] 15

Kaiser-Meyer-Olkin (KMO) Test
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = M)
Overall MSA = 0.7
MSA for each item =

Fe	Mn	Ni	Pb	U	Zn
0.64	0.80	0.71	0.73	0.79	0.64

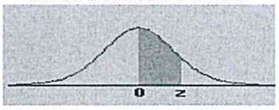
###eigen decomposition
$values ($\lambda$)
[1] 3.063 1.173 0.676 0.509 0.447 ____

$vectors ($\boldsymbol{e}_i$)

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.513	-0.234	0.269	-0.051	0.024	0.779
[2,]	0.436	0.029	-0.365	-0.765	-0.228	-0.195
[3,]	0.286	-0.602	-0.549	0.471	-0.109	-0.145
[4,]	0.323	0.610	-0.046	0.401	-0.600	0.031
[5,]	0.396	0.419	-0.257	0.153	0.758	-0.059
[6,]	0.452	-0.186	0.652	0.078	0.027	-0.574

===========END OF QUESTION PAPER===========

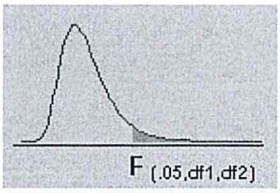
Area between 0 and z



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Table for α=.05

df1



df2

df2\df1	1	2	3	4	5	6	7	8	9	10	12
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882	243.906
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.384	19.396	19.413
3	10.128	9.552	9.277	9.117	9.014	8.941	8.887	8.845	8.812	8.786	8.745
4	7.709	6.944	6.591	6.388	6.256	6.163	6.0942	6.041	5.998	5.964	5.912
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678
6	5.987	5.143	4.757	4.533	4.387	4.284	4.207	4.147	4.099	4.060	3.999
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.676	3.637	3.575
8	5.318	4.459	4.066	3.838	3.688	3.581	3.501	3.438	3.388	3.347	3.284
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.229	3.178	3.137	3.073
10	4.965	4.103	3.708	3.478	3.326	3.217	3.136	3.072	3.020	2.978	2.913
11	4.844	3.982	3.587	3.358	3.204	3.095	3.012	2.948	2.896	2.854	2.788
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.645	2.602	2.534
15	4.543	3.682	3.287	3.056	2.901	2.791	2.707	2.641	2.587	2.544	2.475
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.537	2.494	2.425
17	4.451	3.591	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.441	2.393	2.348	2.278